Construct Validity Studies of the TLAP-R

Xavier Jouve

Cogn-IQ.org
Abstract

The TLAP-R is presumed to yield with the measurement of a second-stratum latent trait chiefly composed of fluid intelligence. Based on rationale of the author, it has been hypothesized that loadings on the fluid intelligence factor and on other cognitive ability factors could vary according to the three-stratum theory. Three samples of participants who took criterion measures were analyzed using factor analysis. Investigations were conducted on groups of 67, 66 and 22 participants who took the following criterion tests: the Advanced Progressive Matrices (APM), the Scholastic Aptitude Test (SAT), the Slosson Intelligence Test-Revised (SIT-R) and a Short Term Memory Retention Test (STMRT). Age of participants has also been used. Results suggested that the measurement yielded by the TLAP-R could be interpreted as a representation of fluid intelligence, with a good predictive validity of crystallized intelligence without much significant aging effect.

*Keywords: TLAP-R, SAT, APM, SIT-R, fluid intelligence, crystallized intelligence, short term memory, factor analysis, aging effect*
Construct Validity Studies of the TLAP-R

The TLAP-R (Jouve, 2013a) is an untimed, non-verbal reasoning test prepared with perceptual material. It consists in 8 matrices, each one including 6 lines of 15 patterns of which 4 have been deleted. The examinee is asked to find the 4 missing patterns. The first version (limited to 5 matrices) of this test has been prepared by Xavier Jouve in mid 2000 for a research project.

The chiefly involved cognitive process mostly depends on fluid intelligence and the test is minimal-knowledge based. The material used to prepare the matrices is numerical but not heavily mathematically based so that the TLAP-R is suitable to assess any person with only basic arithmetic knowledge.

In order to complete the matrices, to find which patterns are missing, the TLAP-R requires inductive reasoning. The task requires figuring out a specific logical and governing rule from a chaotic situation. Spearman (1927) described this process and considered it as the requisite for the eductive part of $g$. However, the TLAP-R showed a strong relationship with crystallized intelligence even without measuring it directly (Jouve, 2013b).

Construct validity is the examination of the theoretical background of the test (Pennington, 2003) by the rationale of the author or through analysis of the empirical data collected. A test is always the empirical representation of the theoretical idea presented by the test designer. If the theoretical ideas of the person who created the test are not correct, construct validity will not be demonstrated. However, if the experiment has been prepared according to a proper rationale, it will demonstrate construct validity.

There are two ways to give an indication of a test’s construct validity. On the one hand, the first method that is issued form differential psychology is an analytic method, which tends to
CONSTRUCT VALIDITY STUDIES OF THE TLAP-R

give an approach based on dimensional measures. In this manner, the construct validity of the assessment is checked if the given measure is similar to the one from a test that is supposed to be tapping the same latent construct. Correlations with equivalent and non-equivalent known, usually published, psychometric tests and factor analytic studies are required to give support for the concrete validity.

On the other hand, the cognitive psychology method for establishing construct validity does not require empirical validity. In fact, the cognitive branch of psychology locates its main interest in the understanding of the functions and processes of the intellect. The specific procedure used by a cognitive psychologist is directed to establishing functional correspondence between the experimental test and another, which is supposed to be an equivalency. If both tests have sufficient equivalent parameters, one can conclude that the experimental test has good theoretical validity.

The rationale of the TLAP-R is above of all a measure of culture-fair reasoning and thus taps a non-verbal type of intellectual ability. We know that most of the non-verbal reasoning abilities are parts of the fluid part of g. In correspondence with this hypothesis, the TLAP-R therefore must be highly loaded on the fluid intelligence factor (2F). However, according to Gustafsson (1984) this factor is quite similar to the general intelligence factor (3G), consequently the TLAP-R must not be totally independent from the crystallized intelligence factor (2C) and other group factors (Carroll, 1993).

Moreover, as the group factors are ranked in the second stratum in the order of their correlation with 3G, the TLAP-R must have higher correlations with tests that are closer to it than tests that are more distant from it. For example, if we suppose that the TLAP-R is ranked in the second stratum close to the 2F factor, its correlations must be higher with a test measuring 2C
than with a test of general memory (2M). Thus its loadings on factors such as fluid and crystallized intelligence, or on a corresponding first stratum vector, must be higher than its loading on any memory related latent trait.

**Method**

**Participants**

Table 1 displays background information for the samples used in this study. All samples were collected as a part of the reliability analyses and the validation of the TLAP-R. The information shows that the first two samples were of equivalent sizes including 67 and 66 persons each, while the third one has been of a smaller size with 22 participants.

In the first cohort of participants, 56.25% attended college. Among them, 10.94% had completed 5 to 8 years at university (Ph.D.’s represented 1.56% of them). As can be seen in Table 1, a slight majority of these people were females (56.71%). This sample was almost equally divided into students (49.21%) and persons having ended their formal education.

The majority of second sample subjects reported having completed or studying towards a college degree: 66.67%, among which, respectively 19.05% and 17.46% mentioned their highest educational level as being a first year, or a second year at university. Post-graduates were 7.94% Masters’ and 3.18% Ph.D.’s. An important part of this sample was composed of students from a variety of academic levels (70.79%).

The third and last group was mainly composed of males (66.67%). To the same extent and fortuitously, the number of students (66.67%) was the largest among occupations given by the participants. 61.91% of subjects were pursuing their college degree or had ended studying
after having completed a college programme. The highest education reported was a Bachelor degree (9.54%). The range of people used for this study was 16 – 24 years of age.

**Measure**

The criterions that have been chosen are widely known and are frequently used in differential and experimental psychology. They have a proper amount of academic literature to show evidence in their behalf.

*Advanced Progressive Matrices (APM).* The APM (Raven, Raven & Court, 1998a) is a nonverbal figurative reasoning test divided into two sets of items. Only the second set has been used for the purposes of this study. It consists in 36 matrices of 3 lines and 3 columns each. The last pattern of the third line is missing, and 6 choices are given to the subject. Only one of these can logically be chosen to complete the matrix.

The APM compose the test of hardest difficulty among the Raven’s Progressive Matrices: the Colored Progressive Matrices (CPM) (Raven *et al.* 1998b) are designed to assess children, the Standard Progressive Matrices (SPM) (Raven *et al.* 2004) aim at appraising average adults and the Advanced Progressive Matrices are dedicated to measure ability in above average adults.

Raven’s tests of matrices have provided with strong evidence over years of being very appropriate measures of fluid intelligence and are usually used as a criterion when the validity of an experimental questionnaire needs to be checked. According to the test manual, the APM scores Cronbach Alpha reliability was .87 in a cohort of 1,015 15 years old German students. Additional data from diverse researches reported split-half coefficients between .83 and .87 (Jensen, Larson & Paul, 1988; Lapsley & Enright, 1979; Paul, 1985; Poortinga, 1972). These are
satisfactory values and indicate a good level of reliability in test scores (Aiken, 2000; Nunnally & Bernstein, 1994). Evidence for close relationships between the APM and other cognitive ability tests are numerous: as a relevant example, we can cite Paul (1985) who reported a correlation of .84 (corrected for restriction of range, \( N = 300 \)) with the Performance IQ of the revised Wechsler Adult Intelligence Scale (WAIS-R) (Wechsler, 1981).

**Scholastic Aptitude Test (SAT).** The SAT (College Board, 2012) is a standardized, three-hour test that measures verbal and mathematical reasoning abilities that students develop over time, both in and out of school. Many colleges and universities use the SAT for admission purposes in order to help predicting successful performance in college. Moreover, the SAT, which was initially developed after an IQ test (Lemann, 1999) and despite of successive revisions, is still strongly correlated to traditional intelligence measures (Jouve, 2010, 2011). Frey and Detterman (2003) reported a correlation of .72 between the recentered SAT and the APM. Additionally, these authors indicated that the pre-recentered SAT correlated from .56 to .82 with intelligence measures, with most correlations being higher than .70 in magnitude. Likewise, a study conducted by Raz, Willerman, Ingmundson, and Hanlon (1983) resulted in a correlation between the SAT and the Culture Fair Intelligence Test (CFIT) (Cattell, Krug, & Barton, 1973) of .81.

The post recentering SAT, in use between 1995 and 2005 and from which the scores have been utilized in this study was divided into two sections, (i) a verbal part with emphasis on critical reading in which vocabulary was tested in the context of reading passages and in analogy and sentence-completion questions and (ii) a mathematical part with emphasis on data interpretation and applied math questions in which calculators were permitted but not required.
Short Term Memory Retention Test (STMRT). The STMRT (Jouve, 2000) was a test developed after Peterson and Peterson paradigm (1959) with the aim of measuring the retention of verbal items in short-term memory. The examinee needed to look at ten lines of four unrelated letters each during two and half minutes, with the instruction of memorizing those quadrigrams. As a matter of interest, the original paradigm used trigrams instead. Once the time was over, and during an equivalent period of time the subject was asked to take a paper-and-pencil task of symbol search as a manner of distracting memorization before having to recall the letters as best possible. The symbol search part of the test was prepared with three given symbols to be encircled each time they were repeated into a 40×40 matrix of look-alike items.

Slosson Intelligence Test – Revised (SIT-R). The SIT-R (Slosson, Nicholson & Hibpsham, 1991; Slosson, 1998) is a test prepared for evaluating crystallized verbal intelligence in native English (children and adults). The 187 SIT-R items are derived from the following cognitive domains: Information, Comprehension, Arithmetic, Similarities and Differences, Vocabulary and Auditory Memory.

Standardized on 2,000 individuals, approximating the contemporary U.S. census, the SIT-R uses a deviation IQ ($SD = 16$). The SIT-R provides a complement to other educational assessments that look at learning ability, readiness or achievement. It has a noticeable scores reliability (Spearman-Brown = .97 in the entire sample) and gives an indication of a wide range of cognitive abilities. The technical manual for the test reports strong correlations (between .80 and .90) with other IQ scales such as the Weschler Intelligence Scale for Children (WISC) (Weschler, 1991) or the Stanford-Binet Intelligence Scale (SBIS) (Thorndike, Hagen & Sattler 1986). Kunen, Overstreet & Salles (1996) reported a correlation of .92 with the recommended abbreviated battery of the SBIS in 191 mentally challenged patients. Furthermore, the pre-
revised SIT was seen to very highly correlates with the SBIS Form LM (Terman & Merrill, 1960) with a Pearson product moment formula of .92 in 724 students (Armstrong, 1971) and showed a range-restriction corrected value of .73 (raw $r = .61$, $N = 98$; Baum, 1979) with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Weschler, 1967).

**Results**

Factor analysis is a common method of examining the patterns of relationships among a set of variables and is a widely used analytic approach in order to investigate the existence and the structure of any latent constructs among a set of items, or tests (Cronbach, 1990; Kamphaus, 2001). For exploring the presence of latent traits among the TLAP-R and the criterion measures, the method of principal components was chosen, with an orthogonal rotation for interpreting the results. The method employed for rotating factors was the varimax method (Kaiser, 1958), which is unquestionably the most frequently and most popular rotation method by far. Varimax rotation simplifies interpretation because each one of the variables is associated with a small number of large loadings and a large number of null to small loadings. As a matter of consequence, extracted factors only represent few variables while they have been only matching few factors.

The first unrotated factor is usually interpreted as $g$ by researchers. However, the nature of any latent trait is determined by the nature of the clusters involved in the study: for example, a principal factor among a set of verbal tasks could be reasonably interpreted as general verbal ability.

**Study 1.** Table 2 shows the results of the principal components analysis (PCA) of the first sample. This sample included 67 scores from the SIT-R and the STMRT along with the TLAP-R. As can be seen, the loadings on the first factor of the TLAP-R and the SIT-R were
somehow equivalent. The figures (.90 and .89 respectively) were very high and indicated that both measures were strongly loaded. At a lower level, the working memory task showed a moderate loading on the principal factor (.44). This factor explained 59.91 percent of the protocol variance. Figure 1 graphically illustrates these findings.

A second PCA has been performed with the same variables, but including also the Age of examinees. Factors eigenvalues were 1.82, 1.19, 0.74 and .25. The use of varimax rotation of the vectors helped exploring the results. These are plotted in Figure 2. In ascending order, loadings of the four variables on the first rotated factor were .94 for the TLAP-R, .88 for the SIT-R, .12 for the STMRT and .08 for the age of participants. The second factor of this analysis was a clear reflect of Age, in which it loaded at .98. Both the TLAP-R and the STMRT were not significantly loaded (-.10 each) by this trait while the SIT-R has been shown a low positive loading (.27). The third latent construct was correlated at .99 with the STMRT and from the fact was identified as a representation of short term memory retention.

Furthermore, the use of a two-dimensional factor structure made sense, and we observed a first factor on which loadings were as follows: .93 for the SIT-R, .87 for the TLAP-R, .35 for the STMRT and .27 for the age. The second factor was characterized by a strong positive loading of the age (.82) while the STMRT has been strongly loaded as well, but negatively (-.70).

Study 2. Plots from a PCA conducted on both the TLAP-R and the 40 minutes timed APM raw scores along with the age of 66 subjects are shown in Figure 3. Values resulted from a varimax rotation. Factors 1 and 3 portrayed the APM and the TLAP-R respectively. Loadings of both tests on their matching factor were .89, and each one loaded .45 on the factor that sketched the other one. The second factor was undoubtedly the age of persons gathered in this sample.
Age loading on this one was perfect (1.00). However, it did not show any significant loading on the two other latent traits (.01 on factor 1 and .03 on factor 3).

Prior to apply varimax rotation, the APM and the TLAP-R appeared equivalently loaded on the principal factor with the same value of .95. Loading of the age of test-takers was .11. Eigenvalues suggested the use of an exploratory two-factorial solution as the value for the first factor was 1.81 (60.28% of the variability), and that of the second factor was 1.00 (33.15% of the variability).

**Study 3.** Another factor analysis has been performed on the scores of 22 individuals, aged 16-24, who took the SAT (recentered) prior to the TLAP-R. Results of confirmatory and exploratory analyses are displayed in Table 3. As shown in Figure 5, unrotated principal factor loadings yield a similar level of proximity for the APM (.92) and the TLAP-R (.90). Although a little less, the verbal reasoning scale of the SAT was also highly loaded (.66) on the principal factor.

Exploratory investigation of the latent constructs among this set of variables revealed a first factor on which the TLAP-R loaded the most (.90) and the mathematical reasoning scale of the SAT loaded significantly (.45). Factor 2 of this solution modeled the verbal reasoning scale of the SAT. In fact, the SAT V accounted for a value of .97.

After adding the age of persons whose scores were collected, a new exploratory factor analysis has been carried out. Eigenvalues were 2.09, 1.15, .56 and .20. Findings for a trifactorial solution are exposed in Figure 4. The TLAP-R loaded at .93 and the SAT M loaded at .92 on the first factor. The SAT V showed only small positive correlation on it (.27). Nevertheless, the SAT V loaded at .96 on the third factor. The second factor matched almost completely the age of testees (.99).
The two-factorial solution gave enlightenment: the first factor represented almost equivalently the SAT M and the TLAP-R which the loading was .93 and .92 respectively. The second factor mimicked very well the age of participants, as age loading was .92. Concerning the SAT V, it showed comparable loadings on both, with .60 and .56.

Discussion

The purposes of this article were to examine the relationships between the TLAP-R and other measures of cognitive ability along with age of participants, with the aim of investigating the construct validity of the TLAP-R. In order to perform this duty, three independent studies using three different samples were conducted. Specifically, two questions were addressed. The main question was about the evidence of convergent validity between the TLAP-R and fluid intelligence factor measured by the APM. The results presented in this article strongly suggest a very close linkage between both, as the TLAP-R and the APM were seen to tap on very closely related domains. In addition to the Pearson product moment correlation between the TLAP-R and the 40 minutes timed APM given by Jouve (2013b), this study provided with even more evidence of the TLAP-R being a measure of 2F.

The second question addressed in this research was the following: is there sufficient evidence to consider the TLAP-R as an indication of a wide range of mental abilities? The results did show evidence of a close relationship with crystallized intelligence as measured by the SIT-R and also with reasoning in scholastic domains, both verbally and non-verbally, as assessed by the college entrance SAT. The TLAP-R showed clear proximity with the mathematical reasoning scale of the SAT, and moderate proximity with verbal reasoning part of this battery. Moreover, findings suggested that the TLAP-R is a relatively fair instrument in regard to the age of
examinees and to a negative effect induced by aging such the decline in short term memory. In other words, where the ability to retained information in short term memory and to recall it efficiently was inversely proportional to the age (both variables being at opposite extrema of the same factor), the TLAP-R remained unimpacted.
References


Jouve, X. (2000). *Test de Rétention en Mémoire à Court Terme (Short Term Memory Retention Test, STMRT)*. Unpublished manuscript.


Table 1

*Descriptive Statistics for Study Samples*

<table>
<thead>
<tr>
<th>Sample</th>
<th>N</th>
<th>Age Range</th>
<th>Age Median</th>
<th>Percentage Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1: SIT-R, STMRT &amp; TLAP-R</td>
<td>67</td>
<td>13-57</td>
<td>23.40</td>
<td>56.71</td>
</tr>
<tr>
<td>Sample 2: APM &amp; TLAP-R</td>
<td>66</td>
<td>14-45</td>
<td>22.11</td>
<td>46.97</td>
</tr>
<tr>
<td>Sample 3: SAT M, SAT V &amp; TLAP-R</td>
<td>22</td>
<td>16-24</td>
<td>20.00</td>
<td>33.33</td>
</tr>
</tbody>
</table>

*Note.* APM = Advanced Progressive Matrices; SAT M = Scholastic Aptitude Test-Recentered Mathematical reasoning scale; SAT V = Scholastic Aptitude Test-Recentered Verbal reasoning scale; SIT-R = Slosson Intelligence Test-Revised; STMRT = Short Term Memory Retention Test.
Table 2

*Principal Components Factor Loadings of the SIT-R, the STMRT and the TLAP-R*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Principal Factor</th>
<th>Factor 1 (GR)</th>
<th>Factor 2 (STMR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIT-R</td>
<td>.90</td>
<td>.38</td>
<td>.10</td>
</tr>
<tr>
<td>STRMT</td>
<td>.44</td>
<td>.07</td>
<td>.97</td>
</tr>
<tr>
<td>TLAP-R</td>
<td>.89</td>
<td>.92</td>
<td>.09</td>
</tr>
</tbody>
</table>

Note. SIT-R = Slosson Intelligence Test-Revised; STMRT = Short Term Memory Retention Test. Principal Factor values are the loadings on the first unrotated factor. Loadings for both Factor 1 and Factor 2 are those following varimax rotation and identified as General Reasoning (GR) and Short Term Memory Retention (STMR).
Table 3

*Principal Components Factor Loadings of the SAT Reasoning Scales and the TLAP-R*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Principal Factor</th>
<th>Factor 1 (NV)</th>
<th>Factor 2 (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT M</td>
<td>.92</td>
<td>.45</td>
<td>.22</td>
</tr>
<tr>
<td>SAT V</td>
<td>.66</td>
<td>.15</td>
<td>.97</td>
</tr>
<tr>
<td>TLAP-R</td>
<td>.90</td>
<td>.90</td>
<td>.17</td>
</tr>
</tbody>
</table>

*Note.* SAT M = Scholastic Aptitude Test-Recentered Mathematical reasoning scale; SAT V = Scholastic Aptitude Test-Recentered Verbal reasoning scale. Principal Factor values are the loadings on the first unrotated factor. Loadings for both Factor 1 and Factor 2 are those following varimax rotation and identified as Nonverbal ability (NV) and Verbal ability (V).
Figure 1. Two-dimensional projection of TLAP-R, STRMT and SIT-R unrotated factor patterns

Note. F1 = Factor 1, i.e. principal factor; F2 = Factor 2, i.e. short term memory; SIT-R = Slosson Intelligence Test-Revised.
Figure 2. Three-dimensional projection of the TLAP-R, SIT-R, STMRT & Age after varimax rotation

*Note.* SIT-R = Slosson Intelligence Test-Revised; STMRT = Short Term Memory Retention Test; Dimension 1 = Problem solving potency; Dimension 2 = Aging effect; Dimension 3 = Short term memory retention.
Figure 3. Three-dimensional projection of TLAP-R, APM & Age after varimax rotation

Note. APM = Advanced Progressive Matrices; Dimension 1 = Speeded figurative problem solving; Dimension 2 = Aging effect; Dimension 3 = Unspeeded figurative problem solving.
Figure 4. Three-dimensional projection of the TLAP-R, SAT M, SAT V & Age after varimax rotation

Note. SAT M = Scholastic Aptitude Test-Recentered Mathematical reasoning scale; SAT V = Scholastic Aptitude Test-Recentered Verbal reasoning scale; Dimension 1 = Nonverbal reasoning; Dimension 2 = Aging effect; Dimension 3 = Verbal reasoning.
Figure 5. Two-dimensional projection of TLAP-R, SAT M and SAT V unrotated factor patterns

Note. F1 = Factor 1, i.e. principal factor; F2 = Factor 2, i.e. verbal reasoning; SAT M = Scholastic Aptitude Test-Recentered Mathematical reasoning scale; SAT V = Scholastic Aptitude Test-Recentered Verbal reasoning scale.